

Unraveling the beautiful complexity of simple lattice model polymers and proteins using Wang-Landau sampling

Thomas Wüst*

*Swiss Federal Research Institute WSL,
Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland*

Ying Wai Li

Center for Simulational Physics, University of Georgia, Athens, Georgia 30602, U.S.A.

David P. Landau

Center for Simulational Physics, University of Georgia, Athens, Georgia 30602, U.S.A.

(Dated: January 16, 2013)

Abstract

We describe a class of “bare bones” models of homopolymers which undergo coil-globule collapse and proteins which fold into their native states in free space or into denatured states when captured by an attractive substrate as the temperature is lowered. We then show how, with the use of a properly chosen trial move set, Wang-Landau Monte Carlo sampling can be used to study the rough free energy landscape and ground (native) states of these intriguingly simple systems and thus elucidate their thermodynamic complexity.

PACS numbers: 87.15.ak, 05.10.Ln, 05.70.Fh, 36.20.Ey

Keywords: lattice homopolymers, lattice proteins, Wang-Landau sampling, protein adsorption

* thomas.wuest@wsl.ch

I. INTRODUCTION

As physics orients itself increasingly towards soft condensed matter and biology, the twin challenges of understanding how linear homopolymers collapse and how proteins fold into their ground states (“native states” in the language of protein biochemistry) are being attacked with ever increasing sophistication. Chain-like polymers are often described by coarse-grained models that reduce the chemical composition to simple “monomers” or “beads” connected in linear fashion by strong chemical bonds [8]. As the temperature is lowered, the non-bonded attraction between monomers cause the polymers to collapse into densely entwined structures [26]. Proteins are polymer chains with complex side groups, all composed of just a few kinds of atoms. In the same spirit, proteins are often represented by simplified linear polymers with **different** monomers in a particular sequence [12]. Protein folding is a complex process which is still poorly understood and which holds the key to unlocking a host of biological and medical secrets. Both of these related types of systems have the complication that the geometric constraints and strong chemical bonds lead to entropic complexity in addition to energetic considerations.

In principle the entire apparatus of statistical mechanics can be brought to bear on the polymer collapse and protein folding problems, but the determination of the thermodynamic properties and ground states of each is still quite challenging. Methods of computer simulation and improvements in computer performance have now reached a level of maturity, however, that they can begin to illuminate the polymer collapse or protein folding processes with good resolution. (Of course, some of the clever techniques that have been developed for magnetic systems, e.g. cluster flipping, do not translate directly to use for the polymer/protein problem.) Hence, the focus of statistical physics is rapidly shifting to the study of biological systems using variations of some of the computational methods that have been well developed for liquids and solids.

In order to permit study of finite temperature behavior, simple polymer/protein models treat the atoms in the chains as classical particles with strong chemical bonds and relatively weak interactions between atoms that are not connected by chemical bonds. Nonetheless, even simple bead-spring models with empirical bonded and non-bonded potentials are still quite difficult to study at low temperature for long chains [26]. Hence, simple lattice models offer the best hope of capturing polymer collapse [12]. Proteins are even more difficult.

Although multiple microscopic, classical “force fields”, i.e. interaction Hamiltonians, have been developed, proteins are sufficiently complex that these parametrized interaction fields are still of rather limited value. In an attempt to make the problem more tractable, Dill [11] introduced a very simplified model which he believed contained the essential features of the protein folding problem without including all of the complexity. In Dill’s approach, proteins are modeled as lattice polymers which contain only two kinds of monomers: hydrophobic (H) and polar (P). The goal was to provide a model that was amenable to quantitative analysis and thus, it served as a basis for - and inspired - computational studies of various biophysical systems [9]. However, it turns out that even Dill’s simplified protein model is very challenging for all but the shortest sequences. Finding the ground states for an HP sequence is proven to be an NP-hard problem [7].

In contrast to molecular dynamics, which is generally restricted to time scales of a few tens or hundreds of nano-seconds, clever Monte Carlo algorithms, which do not follow a physical time scale, can be used to explore behavior at much longer time scales [16, 21]. (Of course, for lattice models molecular dynamics does not work at all.) The “workhorse” Monte Carlo method in statistical physics, developed almost 60 years ago, has been the Metropolis importance sampling algorithm [25]; but more recently new, efficient algorithms have begun to play a role in allowing simulation to achieve the resolution which is needed to accurately locate and characterize “interesting” thermodynamic behavior [6, 8, 13, 18, 21]. Traditional methods show quite long time correlations and require excessively long runs to obtain precise results. As we shall see, for certain lengths homopolymers may have complicated behavior as the temperature is lowered. These difficulties also hold true for protein models that, as expected, have complex free energy landscapes with energy barriers between states of low free energy [1]. Consequently, “standard” Monte Carlo methods become easily “trapped” in metastable states from which they have difficulty escaping.

Here we review a relatively new, general, efficient Monte Carlo algorithm (known as “Wang-Landau sampling” [31]) that offers substantial advantages over other approaches [28]. We will then describe the application of this algorithm to the study of lattice polymers of different chain length and then to the HP model for several “characteristic” lattice protein sequences. One step further in complexity, we will also present our work on the problem of protein adsorption where we investigate HP model proteins interacting with an attractive surface.

There have been a substantial number of other studies performed with different simulation methods, and we shall try to place our results in perspective.

II. THE LATTICE HOMOPOLYMER AND HP PROTEIN MODELS

Our lattice homopolymer model consists of a sequence of identical monomers on sites of a periodic lattice, and these monomers are connected together by unbreakable nearest neighbor bonds. The weak interactions between monomers which are chemically separated but physically nearby are modeled by an attractive nearest neighbor potential; otherwise they do not interact. The HP model is a simple, prototypical lattice protein model consisting of only two types of monomers, hydrophobic (H) and polar (P), in a sequence chosen so as to mimic a real protein [11]. The chemical bonds between monomers cannot break, and **non-bonded** interactions are restricted to an attractive coupling, ε_{HH} , between hydrophobic monomers that occupy nearest-neighbor sites. Although the model is obviously over-simplified, it is intended to capture the hydrophobic character, which is considered as the “driving force” for protein folding [19]. Moreover, in spite of the simplicity of the model, for long chain lengths it is often very challenging to determine the lowest possible energy (native state) for the folded chain.

In the presence of an attractive surface, the homopolymer and protein models can be easily extended to include the interaction between the monomers and the surface. (Here, we shall confine ourselves to the protein case. For an overview of lattice homopolymers interacting with surfaces, see Ref. [3].) In the simplest case for proteins the surface attracts both types of monomers with strength ε_s and is introduced to the 3-dimensional lattice as an xy -plane placed at $z = 0$. A second, non-attractive wall is placed at $z = h_w$ to keep the HP chain from diffusing far away from the bottom surface, but h_w is still sufficiently large that the protein can still be fully extended while in contact with the surface. Periodic boundaries are used for the x and y directions. The Hamiltonian of the model is then represented as [4]:

$$\mathcal{H} = -\varepsilon_{HH}n_{HH} - \varepsilon_s n_s, \quad (1)$$

n_{HH} being the number of H-H interacting pairs and n_s being the number of monomers adjacent to the bottom surface.

III. THE WANG-LANDAU SAMPLING ALGORITHM

The shortcomings of traditional Monte Carlo algorithms led to the development of a different, iterative approach to Monte Carlo sampling which executes a random walk in energy space rather than in probability space. This algorithm is now widely referred to as “Wang-Landau sampling” and has already been described in detail for interacting magnetic spin systems [22]. The same basic approach applies to lattice polymers/proteins, but suitable trial moves must be implemented for the method to work efficiently [32]. Unlike the Metropolis Monte Carlo method, Wang-Landau sampling does not depend upon the system temperature and focuses instead on the density of states $g(E)$. If we perform a random walk in energy space by moving monomers, and the probability to visit a given energy level E is proportional to the reciprocal of $g(E)$, then a flat histogram of energies that have been visited is generated. This is done iteratively by modifying the estimated $g(E)$ to produce a “flat” histogram of energies that have been visited (over the allowed range of energy) while making $g(E)$ converge to the true value. During each succeeding step of the random walk, $g(E)$ is modified and the updated value is used to continue the random walk in energy space.

Initially, $g(E)$ can be set to $g(E) = 1$ for all possible energies E . Then a random walk in energy space is begun. If E_1 and E_2 are energies before and after a monomer is moved, the transition probability from energy level E_1 to E_2 is

$$p(E_1 \rightarrow E_2) = \min \left(\frac{g(E_1)}{g(E_2)}, 1 \right). \quad (2)$$

Each time an energy level E is visited, we multiply the existing value of $g(E)$ by a “modification factor” $f > 1$. (As a practical matter it is preferable to work with the logarithm, i.e. $\ln(g(E)) \rightarrow \ln(g(E)) + \ln(f)$, so that all possible values of $g(E)$ will fit into double precision numbers.) If a trial move is rejected, the system stays at the same energy and the existing density of states is multiplied by the modification factor. While multiple choices are possible, a reasonable value of the initial modification factor is $f = f_0 = e^1 \simeq 2.71828 \dots$. During the random walk, we also accumulate a histogram $H(E)$, i.e. the number of visits at each energy level E , in energy space. When the histogram is sufficiently “flat” over the entire energy range of the random walk, $g(E)$ converges to the true value with an accuracy proportional to that modification factor $\ln(f)$. We then reduce the modification factor using a function like $f_1 = \sqrt{f_0}$, reset the histogram, and begin the next iteration of the random walk. This process continues until the histogram is “flat” again and we then reduce the modification

factor $f_{i+1} = \sqrt{f_i}$ and restart. The process is halted when the modification factor is smaller than some predefined value. (We determined that $f_{final} = \exp(10^{-8}) \simeq 1.00000001$ was a good choice for the models considered here.) Note that the phrase “flat histogram” refers to a histogram $H(E)$ for which all entries are not less than $x\%$ of the average histogram $\langle H(E) \rangle$, where $x\%$ depends upon the desired accuracy of the density of states.

During the early iterations the algorithm does not satisfy the detailed balance condition, since $g(E)$ is constantly modified; however, after many iterations, $g(E)$ converges to the true value as the modification factor approaches 1. If $p(E_1 \rightarrow E_2)$ is the transition probability from the energy level E_1 to level E_2 , from equation (2), the ratio of the transition probabilities from E_1 to E_2 and from E_2 to E_1 can be calculated very easily as

$$\frac{p(E_1 \rightarrow E_2)}{p(E_2 \rightarrow E_1)} = \frac{g(E_1)}{g(E_2)} \quad (3)$$

where $g(E)$ is the density of states. Thus, the random walk algorithm satisfies detailed balance. The proof of convergence, i.e. finding of the correct $g(E)$, of the Wang-Landau algorithm has been given rather in view of an optimization procedure than by means of Markov chain Monte Carlo arguments[34]. In this respect the condition of detailed balance does not bear much importance (note that thermodynamic quantities are obtained only indirectly from $g(E)$). Nonetheless, for certain trial move sets, it is important to correct Eq. (2) for unequal move ratios in order to fulfill detailed balance and to avoid a possible bias in the final estimate of $g(E)$ [32].

A. The importance of trial move sets

Simple trial moves that allow single monomer, or only a few neighboring monomers, to move, (e.g. end-bond flips, kink jumps or crankshaft moves, as shown in Fig. 1) have been developed and used for many years for the study of geometric properties of lattice polymers. These simple moves by themselves do not produce an ergodic process, although the inclusion of “pivot” moves restores ergodicity [29]. This does not solve all the problems, however, since the acceptance rate of “pivot” moves becomes extremely low for compact polymers. For this reason, additional trial moves need to be implemented, regardless of the Monte Carlo algorithm used.

We have adopted two types of trial moves in our Wang-Landau sampling simulations and

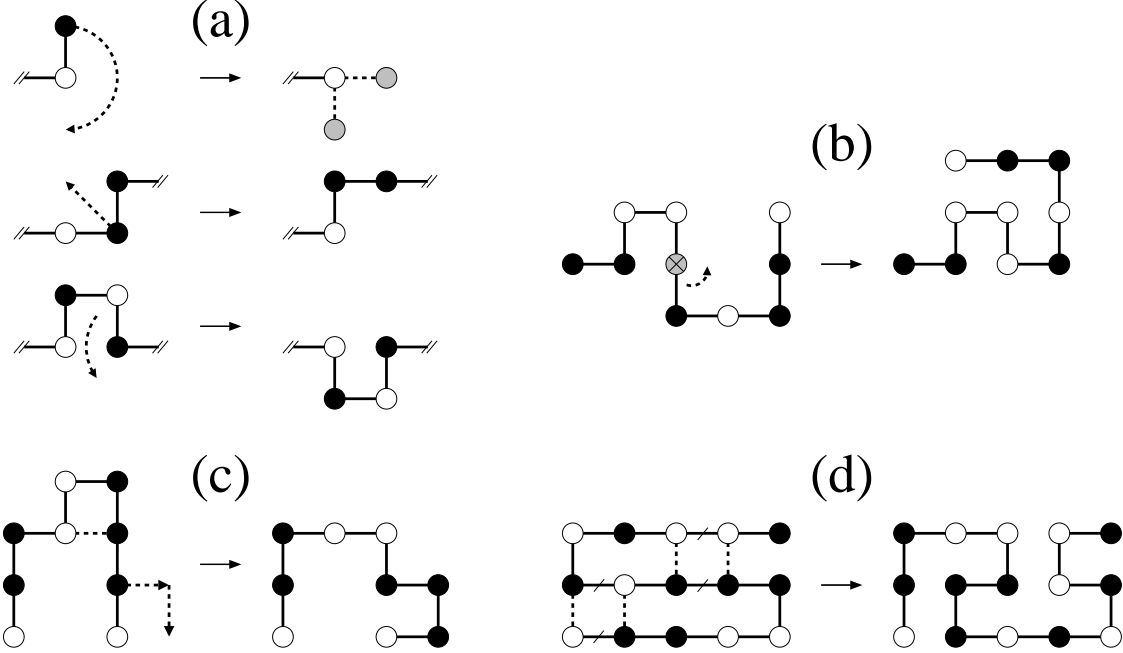


FIG. 1. Examples of trial moves typically applied in Monte Carlo simulations of lattice polymers and proteins (here shown for a 2D sc lattice only). (a) End-bond flip, kink jump and crankshaft move; (b) pivot move; (c) pull move; (d) bond-rebridging move. These types of trial moves can be adopted to systems of any dimension > 2 .

determined that their use proves to be extremely effective at accelerating convergence [32]. The first kind of trial move, “pull moves”, effectively threads all of, or a portion of, the polymer along the existing, entangled polymer structure [23]. A second kind of trial move, “bond-rebridging”, does not actually move any monomers but cuts bonds on opposite sides of an elementary plaquette and redraws them on the other two sides to change the polymer conformation dramatically in a single trial move [10]. Since the monomers do not move, bond-rebridging alone is not ergodic, but we shall combine pull moves with bond-rebridging and ergodicity is no longer a problem. Fig. 1 shows examples of “traditional” trial moves as well as of the ones applied in our Monte Carlo simulations (pull and bond-rebridging moves).

IV. THREE DIMENSIONAL LATTICE HOMOPOLYMERS

Since all monomers are equivalent, obtaining results for multiple values of the monomer number N will enable us to attempt to determine the systematic behavior of the polymer

collapse as the length diverges. However, the length dependence is not at all simple. Those homopolymers of length $N = L^3$, where L is an integer, are special in that their ground state is a perfect cube of monomers. (Note, because of the constraints due to the bond connectivity, the attainment of this cubic overall structure involves quite complicated configurations.) If the polymer is slightly longer or shorter than a “magic number”, the behavior may change dramatically because of the extra degrees of freedom for those monomers that lie outside a perfect cube. For now, we will concentrate on homopolymers whose length is a “magic number” so that we can compare behaviors as the length increases. The specific heat demonstrates the complexity of the behavior as the temperature is reduced. First, as shown in Fig. 2 for $N = 27$ a high temperature shoulder with a low temperature pronounced peak is found. For $N = 64$ a rounded maximum appears to signal the polymer collapse into a globule. At much lower temperature, however, another, sharper peak announces the rearrangement of bonds to allow the polymer to assume a configuration that is very close to the ground state, i.e. there is kind of a “freezing transition”. With increasing chain length the upper peak grows in magnitude monotonically and shifts, systematically to higher temperature. In contrast, as the chain length increases, the lower temperature peak shifts downwards in temperature and decreases in magnitude. This shows clearly that in the present model (simple cubic lattice homopolymer model) no symmetry breaking is taking place between a dense liquid and a crystal state (both have simple cubic symmetry). The origin and sharpness of the low temperature peak is rather the result of geometrical surface reconstructions which involve the simultaneous displacement of many monomers. To provide some insight into the homopolymer behavior, in the lower part of the figure we show “snapshots” of typical states of an $N = 125$ homopolymer at different temperatures above and below the two peaks. In addition, there is interesting structure developing just above this sharp, low temperature peak. At first a small, broadened peak appears, but it then flattens out and the specific heat then begins to show small oscillations which appear to be larger than the statistical error bars. According to “snapshots” of the homopolymer structure, these small oscillations seem to result from shifts of compact portions of the ground state cube to create distorted surfaces. Clearly, high resolution results for still longer chains would be highly desirable to illuminate the actual behavior at, and just above, the low temperature transition. (To be published.)

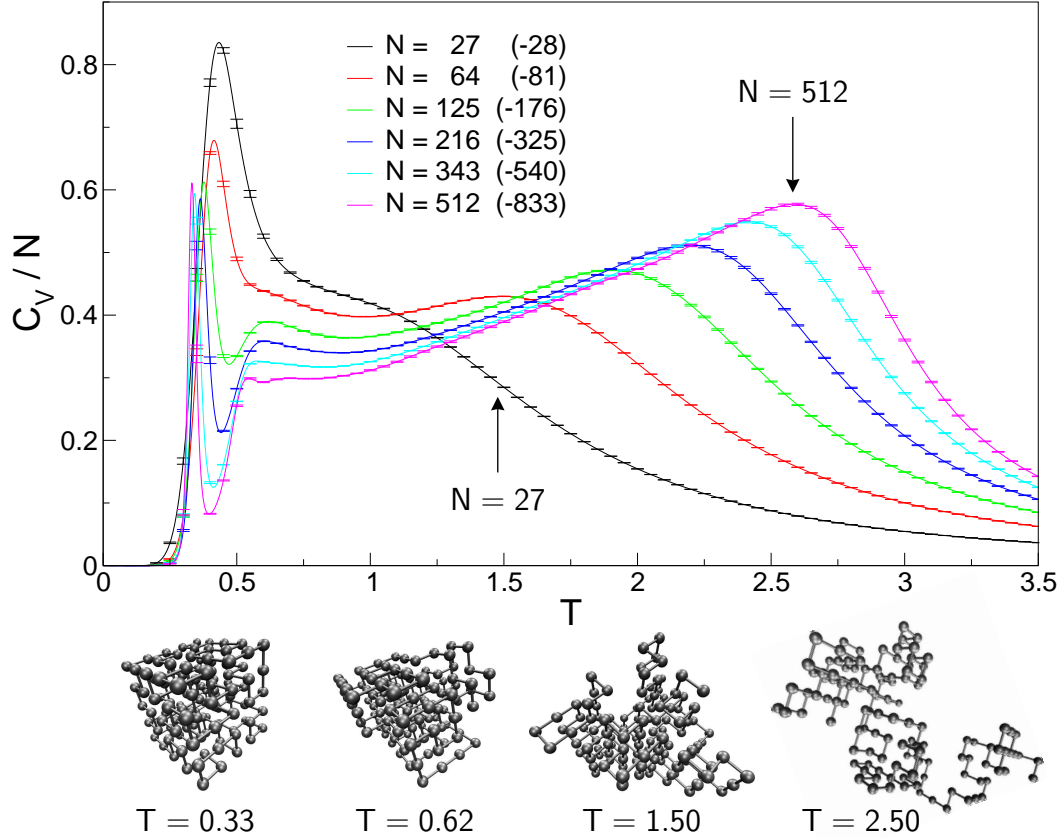


FIG. 2. (Color online) Specific heat as a function of temperature for lattice homopolymers of various chain lengths N on the simple cubic lattice. **Bottom rows:** Representative structures at specific temperatures for $N = 125$. (From Wüst and Landau [32])

V. STRUCTURAL AND THERMODYNAMIC PROPERTIES OF SOME HP MODEL PROTEINS

A. Free HP proteins

Unlike the simple, lattice homopolymers, the monomers of HP lattice proteins are not all equivalent. Each protein has a specific sequence of H 's and P 's, and there is no way to extrapolate the behavior to that of an infinite length equivalent. A number of specific choices of length and sequence have served as “benchmarks” in the literature; and, for pedagogical reasons, we shall emphasize only a few three dimensional systems in the current treatment. The sequence 3D103 (PPHHPPPPPHHPPHHPPHPPPPPPHPP-PHHPHHPPPPPPHPPHPPHPPPPPHHHPPPHHPHPPPPPHHPPPPHHHHHPH-PPPPPPPHHHHHPPHPP) is a benchmark sequence whose native state has proven to be

particularly difficult to determine, so we will now concentrate our discussion on this system. (We note, however, that a great deal of work has also been carried out on two dimensional sequences as well.)

For very short HP sequences, the native state is accessible, in spite of the very large number of distinct configurations that result because of the bond connectivity constraints. For some protein models the native state is degenerate, and in some cases there are multiple sequences of the same length that give different results for both native state energies and temperature dependent behavior. For example, Bachmann and Janke [2] examined several different 14mer HP sequences by exact enumeration and found noticeably different specific heat curves. For longer sequences, however, the problem has proven to be much more challenging. As an example of the relative success that various computational methods have had in finding native state energies for current benchmark HP sequences, in Table I we show results from diverse numerical methods for the lowest energies found. The table includes results not only for 3D103 but also for benchmark sequences in two dimensions with almost the same number of monomers. We note that 2D100 actually represents two different benchmarks with the same number of monomers but different sequences. Furthermore, the ground state energies of 2D100a and 2D100b are **not** the same. In addition to values from Wang-Landau sampling (WLS), we also show results from several other sophisticated approaches: equi-energy sampler (EES) [20], multicanonical chain growth(MCCG) [2, 27], multi-overlap ensemble (MSOE) [17], fragment-regrowth Monte Carlo (FRESS) [33], and pruned-enriched Rosenbluth sampling (PERM) [15]. Note that for “fairness” reasons only methods which are capable (at least principally, see [33]) to sample thermodynamic quantities or even the entire DOS are included in this comparison, whereas methods solely targeting the search for the optimal ground state are excluded (such as e.g. the very efficient constraint-based approach for protein structure prediction [5]).

The characteristic behavior of the specific heat as determined by Wang-Landau sampling is shown for sequence 3D103 in Fig. 3. At high temperature the protein is distended, but as the temperature is lowered there is a rounded peak that signals the collapse of the protein into a compact object, but the acquisition of the native state only begins at a lower temperature where a shoulder occurs in the specific heat. The radius of gyration, also plotted in Fig. 3, decreases substantially in the vicinity of the specific heat peak, and sample configurations (shown in the bottom portion of the figure) confirm both the protein collapse as well as the

TABLE I. Minimum energies found by several methods for benchmark HP sequences in 2D and 3D. The first column names the sequence (dimension and length), see [33]. For abbreviations, see text.

Seq.	WLS	EES	MCCG	MSOE	FRESS ^a	PERM
2D100a	-48	-48	–	-47	-48	-48
2D100b	-50	-49	–	-50 ^b	-50	-50
3D88	-72	–	–	–	-72	-69
3D103	-58	–	-56	–	-57	-55
3D124	-75	–	–	–	-75	-71
3D136	-83	–	–	–	-83	-80

^a Ground state search only.

^b DOS not attained.

lack of a native state until below the temperature at which the shoulder appears. Attempts to determine the native state energy have an extensive history and estimates were slowly descending to lower values with time. The lowest energy obtained so far is $E_{min} = -58$ which was obtained from Wang-Landau sampling, but in spite of the sophistication of the method, it has not been possible to get a reliable estimate of the relative probability of the native state and first excited state.

This native state has a well formed hydrophobic core but no obvious symmetry to distinguish it from the first excited state. The Jaccard index measures the overlap between the configuration of an HP chain and one of the native states [14]. For 3D103 the Jaccard index increases as the temperature is lowered, but it saturates at $T \approx 0.1$ (see Fig. 3) reflecting the degeneracy of the native state.

Fig. 4 shows the specific heat as a function of temperature for the three-dimensional benchmark HP sequences listed in Table I. Comparison among these curves shows clearly their common underlying thermodynamic behavior; that is, the occurrence of a protein collapse (C_V peak at higher temperatures) and another “folding transition” to the native state(s) at low temperatures. However, the characteristics of the later transition shows significant sequence and length dependences ranging from a sharp peak (3D88) to a soft shoulder (3D124) only. This is particularly remarkable in consideration that for all four

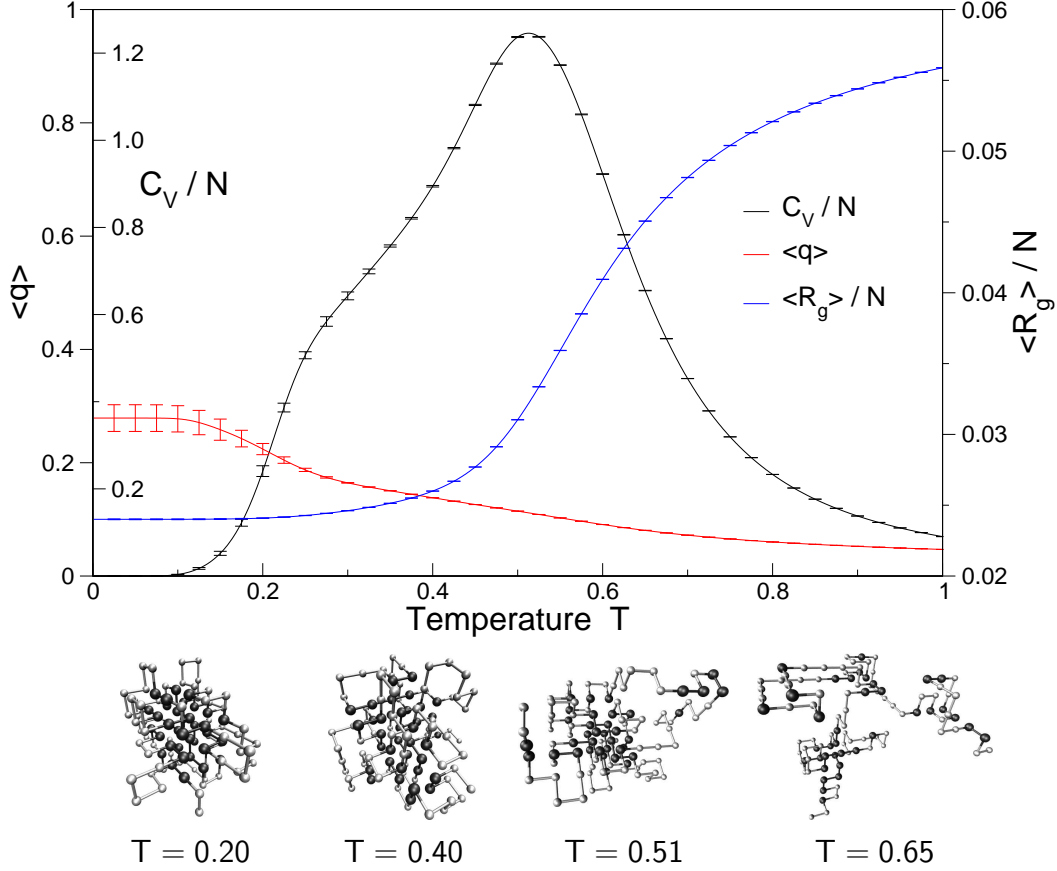


FIG. 3. (Color online) Specific heat C_V/N , mean radius of gyration $\langle R_g \rangle/N$ (N , chain length), and mean Jaccard index $\langle q \rangle$ as a function of temperature for HP sequence 3D103. (From Wüst and Landau [32]) **Bottom rows:** Typical conformations at specific temperatures. The dark spheres represent H -monomers and the light spheres are the P -monomers.

sequences the sampling of low temperature range is equally difficult.

B. Thermodynamic properties of the HP model near an attractive surface

To demonstrate the richness of the “transition” behaviors of the HP chain in the vicinity of an attractive surface, we examined an HP sequence with 36 beads (PPPHHPPHHPPPPPH-HHHHHHPPHHPPPPHHPHPP) and a surface which attracts both H - and P -monomers. We focus on two cases: a weak attractive surface ($\varepsilon_{HH} = 12, \varepsilon_s = 1$) and a strong attractive surface ($\varepsilon_{HH} = 1, \varepsilon_s = 2$), to compare the effects of different surface strengths on the “transition” behaviors. Technical details have been discussed in Ref. [24]. Similar studies can be found in Refs. [4, 30].

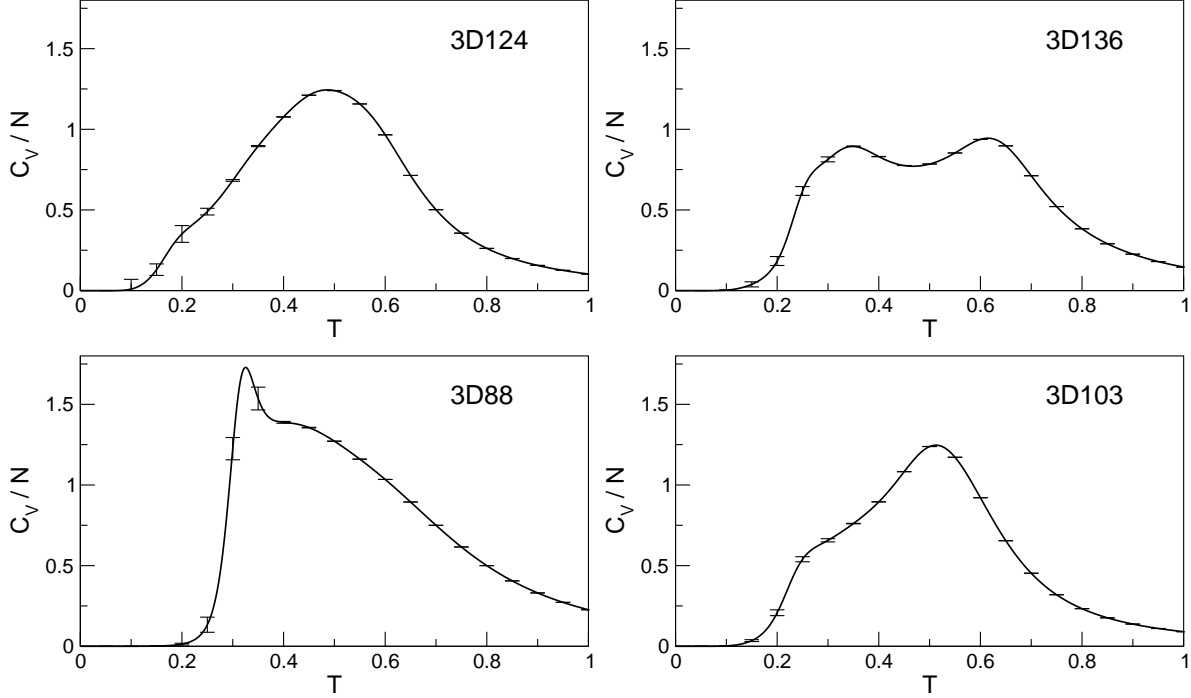


FIG. 4. Specific heat C_V/N as a function of temperature (T) for the three-dimensional benchmark HP sequences listed in Table I. Due to the difficulty to sample the low energies, for sequences 3D103 and 3D136, the DOS has been attained only to energy levels -57 and -81 , respectively. Error bars have been obtained by means of bootstrap resampling.

As in the case of free proteins, the use of “pull moves” and “bond rebridging” turn out to be essential for accessing low energy states relatively rapidly. The top portion of Fig. 5 shows a typical specific heat of an HP chain interacting with a weakly attractive surface, as compared to that of the free protein sequence without the surface. In the former case, three peaks are observed to signal three “transition” stages: the largest peak at $T/\varepsilon_{HH} \approx 0.5$ represents the same coil-globule transition as found in free space. (The comparison of the specific heat from the two cases in Fig. 5 demonstrates this very clearly.) During this stage, the HP polymer transforms from an extended chain-like structure to a compact, but desorbed, globule. Typical structures are much the same as the collapsed states of the free chain in the absence of the surface.

The middle peak at $T/\varepsilon_{HH} \approx 0.18$ signals protein adsorption during which the compact HP globule “docks” at the surface with the hydrophobic core remaining intact. The globule spans several layers vertically and the total energy of the system is lowered slightly due to

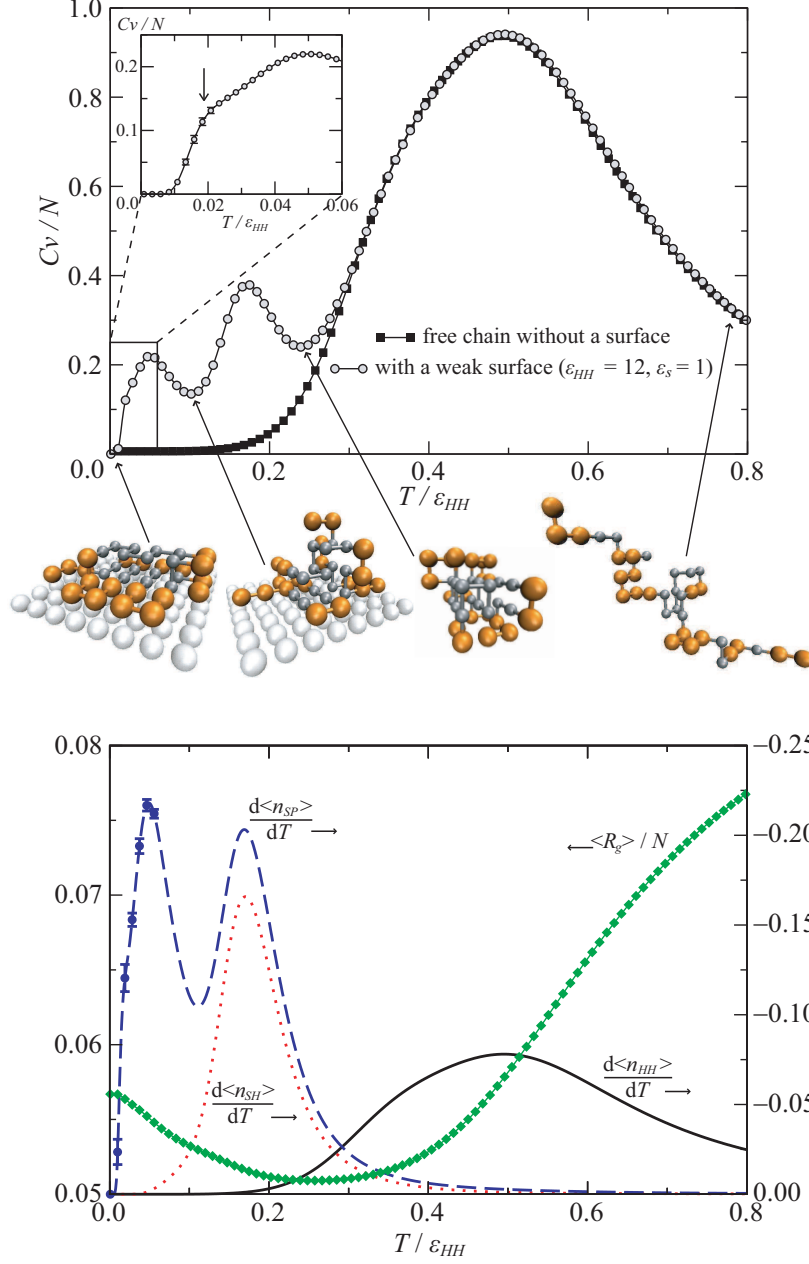


FIG. 5. (Color online) **Top:** Specific heat of HP sequence 36mer interacting with a weak attractive surface ($\epsilon_{HH} = 12, \epsilon_s = 1$) and without the presence of the surface. Typical configurations are shown for several different temperatures. The small spheres represent H -monomers and the large spheres are P -monomers. The faint spheres at the bottom for the two lowest temperatures show the surface. (From Li, Wüst and Landau [24]) **Bottom:** Radius of gyration and thermal derivatives of the numbers of H - H contacts as well as surface contacts. Horizontal arrows besides the legends indicate the scales that the quantities are using. For both graphs, error bars smaller than the data points are not shown.

contact with the surface.

Further decrease in temperature brings the system to the third transition at $T/\varepsilon_{HH} \approx 0.05$ where the system maximizes the number of surface interactions without sacrificing an intact, energetically minimized hydrophobic core. Forming H - H contacts is energetically much more favorable than forming surface contacts with a large value of ε_{HH} . The last interesting feature in the specific heat is a subtle shoulder at $T/\varepsilon_{HH} \approx 0.02$. Typical states at $T/\varepsilon_{HH} = 0$ and $T/\varepsilon_{HH} \approx 0.02$ suggests this is due to the difference in shapes of hydrophobic cores and the number of P -monomer-surface contacts available at these temperatures. At $T/\varepsilon_{HH} = 0$ where there are just ground states, only rectangular H -cores allow a maximum number of P -surface contacts that minimizes the energy.

The temperature dependence of the structural properties is shown in the bottom portion of Fig. 5. At the temperature for which the specific heat shows the largest peak, the radius of gyration is rapidly decreasing and the number of H - H contacts is rapidly increasing (which results in a peak in its thermal derivative) on cooling. Both behaviors support the idea of a “coil-globule” transformation. At the temperature where the middle specific heat peak occurs, the thermal derivative of the number of H -surface contacts also shows a peak. The thermal derivative of the number of P -surface contacts has maxima at the same temperatures as both of the low temperature specific heat peaks. It also shows a weak shoulder at the same temperature as the very low temperature specific heat shoulder ($T/\varepsilon_{HH} \approx 0.02$). The radius of gyration also increases at low temperature because the protein is flattening out on the surface and becoming rather two dimensional in shape.

However, if the surface is strongly attractive, the transition behavior is quite different from the previous case as shown in Fig. 6. For the surface strength we have chosen, only two peaks with a weak bump in between are seen in the specific heat. The peak at $T/\varepsilon_{HH} \approx 3.9$ indicates a protein adsorption transition, and the peak at $T/\varepsilon_{HH} \approx 0.35$ corresponds to the protein collapse transition that takes place completely on the surface. The bump at $T/\varepsilon_{HH} \approx 2.0$ then signals the flattening stage.

Understanding this in conjunction with the structural properties, it is obvious to see that at high temperature where the specific heat shows a peak, the extended HP chain first adsorbs on the surface. This is in agreement with the peaks in the derivatives of the number of surface contacts at the same temperature. In the temperature range $T/\varepsilon_{HH} \approx 3$ down to $T/\varepsilon_{HH} \approx 1$, the partly adsorbed chain flattens itself out until all monomers come in contact

with the surface. Again, the specific heat and the derivatives of the number of surface contacts echo each other by revealing convex bumps at the same temperature. The radius of gyration increases rapidly in this temperature range as the structure becomes more and more planar. However, the chain remains extended without forming many $H-H$ contacts during this stage as no signal is found from the derivative of the number of $H-H$ contacts. As a third step the fully adsorbed, two-dimensional chain undergoes a “collapse” transition at $T/\varepsilon_{HH} \approx 0.35$ to maximize the number of $H-H$ contacts before adopting a film-like, compact structure. This is clearly signaled by the sharp peaks in the specific heat and the derivative of the number of $H-H$ contacts at that temperature.

As shown in the two examples here, structural transitions may occur in different orders for different surface attractive strengths when temperature changes. Some transitions may not give distinct signals in specific heat but merely in particular structural parameters. It is thus essential to analyze specific heat together with structural quantities to identify various “phases”. With longer HP chains, our interpretation of selected structural parameters suggests that the hierarchies of structural phase transitions for this model can be generalized into a few categories. (To be published.)

VI. DISCUSSION AND CONCLUSION

Homopolymer collapse and protein folding remain challenging problems, even for the prototypical lattice models discussed in this article. We described the implementation of an efficient Monte Carlo algorithm (Wang-Landau sampling) to calculate the density of states directly for both homopolymers and HP models with different lengths and monomer sequences. In both cases there were a combination of energetic and entropic barriers that made the convergence difficult. These problems could be largely overcome with Wang-Landau sampling with a combination of pull moves and bond-rebridging moves and with the choice of a relatively stringent final modification factor. Examination of a combination of thermodynamic and structural properties, along with visualization of typical configurations, could be interpreted coherently to explain the physical behavior that was taking place, both with and without the presence of a surface. Other HP sequences remain to be studied, and the low temperature behavior of even longer homopolymers requires careful examination.

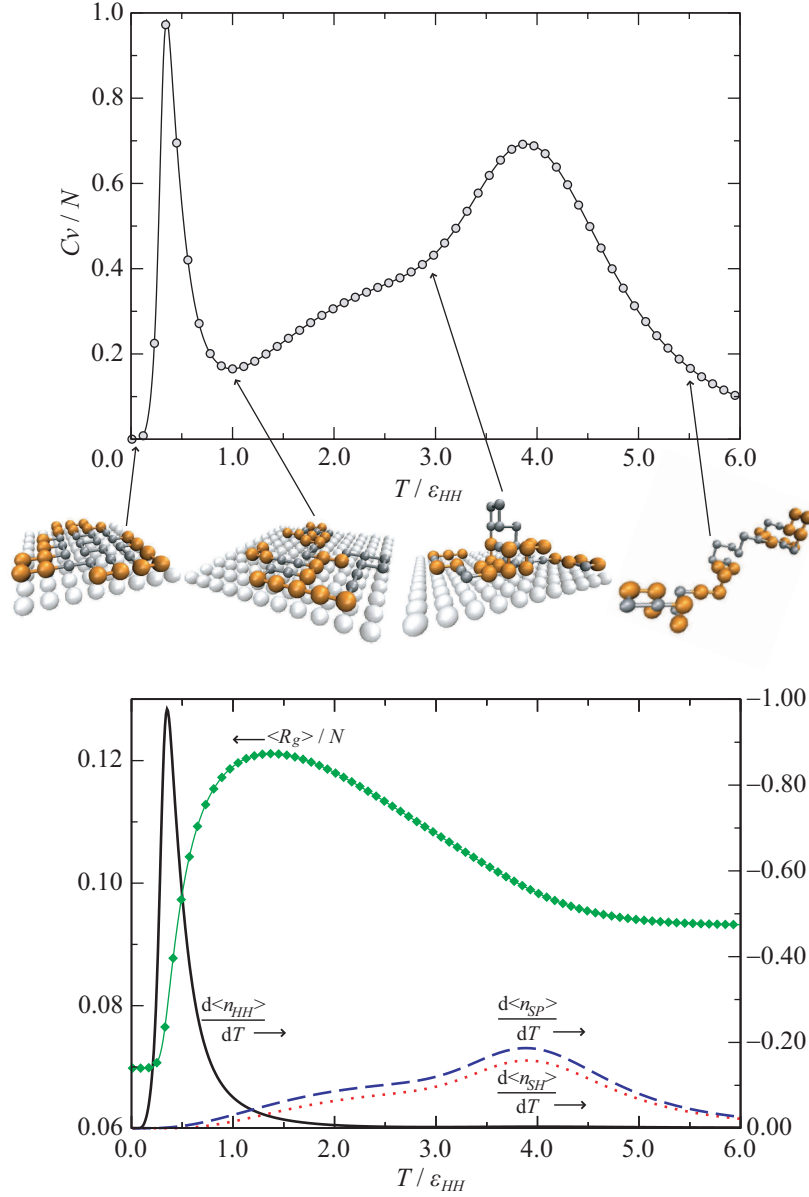


FIG. 6. (Color online) **Top:** Specific heat of HP sequence 36mer interacting with a weak attractive surface ($\epsilon_{HH} = 1, \epsilon_s = 2$). Typical configurations are shown for several different temperatures. The small spheres represent H -monomers and the large spheres are P -monomers. The faint spheres at the bottom for the two lowest temperatures show the surface. **Bottom:** Radius of gyration and thermal derivatives of the numbers of H - H contacts as well as surface contacts. Horizontal arrows besides the legends indicate the scales that the quantities are using. For both graphs, error bars are not shown as all are smaller than the data points.

ACKNOWLEDGMENTS

We thank M. Bachmann and W. Janke for illuminating discussions. The research project is supported by the National Science Foundation under Grant No. DMR-0810223.

- [1] Anfinsen, C.B.: Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973); Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G.: Funnels, pathways and the energy landscape of protein folding: A synthesis. *Proteins* **21**, 167-195 (1995); Onuchic, J.N., Luthey-Schulten, Z., Wolynes, P.G.: Theory of protein folding: The energy landscape perspective. *Ann. Rev. Phys. Chem.* **48**, 545-600 (1997)
- [2] Bachmann, M., Janke, W.: Multicanonical chain-growth algorithm. *Phys. Rev. Lett.* **91**, 208105 (2003); Thermodynamics of lattice heteropolymers. *J. Chem. Phys.* **120**, 6779-6791 (2004)
- [3] Bachmann, M., Janke, W.: Conformational transitions of nongrafted polymers near an absorbing substrate. *Phys. Rev. Lett.* **95**, 058102 (2005); Substrate adhesion of a nongrafted flexible polymer in a cavity. *Phys. Rev. E* **73**, 041802 (2006); Luettmer-Strathmann, J., Rampf, F., Paul, W., Binder, K.: Transitions of tethered polymer chains: A simulation study with the bond fluctuation lattice model. *J. Chem. Phys.* **128**, 064903 (2008)
- [4] Bachmann, M., Janke, W.: Substrate specificity of peptide adsorption: A model study. *Phys. Rev. E* **73**, 020901(R) (2006)
- [5] Backofen, R., Will, S.: A constraint-based approach to fast and exact structure prediction in three-dimensional protein models. *Constraints* **11**, 5-30 (2006)
- [6] Berg, B.A., Neuhaus, T.: Multicanonical algorithms for first order phase transitions. *Phys. Lett. B* **267**, 249-253 (1991); Berg, B.A., Neuhaus, T.: Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* **68**, 9-12 (1992)
- [7] Berger, B., Leighton, T.: Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete. *J. Comput. Biol.* **5**, 27-40 (1998); Crescenzi, P., Goldman, D., Papadimitriou, C., Piccolboni, A., Yannakakis, M.: On the complexity of protein folding. *J. Comput. Biol.* **5**, 423-465 (1998)

- [8] Binder, K., Paul, W.: Recent developments in Monte Carlo simulations of lattice models for polymer systems. *Macromolecules* **41**, 4537-4550 (2008)
- [9] See, e.g., Bonaccini, R., Seno, F.: Simple model to study insertion of a protein into a membrane. *Phys. Rev. E* **60**, 7290-7298 (1999); Harrison, P.M., Chan, H.S., Prusiner, S.B., Cohen, F.E.: Thermodynamics of model prions and its implications for the problem of prion protein folding. *J. Mol. Biol.* **286**, 593-606 (1999); Castells, V., Yang, S., Van Tassel, P.R.: Surface-induced conformational changes in lattice model proteins by Monte Carlo simulation. *Phys. Rev. E* **65**, 031912 (2002); Ping, G., Yuan, J.M., Vallieres, M., Dong, H., Sun, Z., Wei, Y., Li, F.Y., Lin, S.H.: Effects of confinement on protein folding and protein stability. *J. Chem. Phys.* **118**, 8042-8048 (2003); Cellmer, T., Bratko, D., Prausnitz, J.M., Blanch, H.: Protein-folding landscapes in multichain systems. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11692-11697 (2005); Zhang, L., Lu, D., Liu, Z.: How native proteins aggregate in solution: a dynamic Monte Carlo simulation. *Biophys. Chem.* **133**, 71-80 (2008)
- [10] Deutsch, J.M.: Long range moves for high density polymer simulations. *J. Chem. Phys.* **106**, 8849-8854 (1997)
- [11] Dill, K.A.: Theory for the folding and stability of globular proteins. *Biochemistry* **24**, 1501-1509 (1985); Lau, K.F., Dill, K.A.: A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* **22**, 3986-3997 (1989)
- [12] Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S.: Principles of protein folding - A perspective from simple exact models. *Protein Sci.* **4**, 561-602 (1995); Dill, K.A.: Polymer principles and protein folding. *Protein Sci.* **8**, 1166-1180 (1999); Kolinski, A., Skolnick, J.: Reduced models of proteins and their applications. *Polymer* **45**, 511-524 (2004)
- [13] Ferrenberg, A.M., Swendsen, R.H.: New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* **61**, 2635-2638 (1988); Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **63**, 1195-1198 (1989)
- [14] Fraser, R., Glasgow, J.I.: A demonstration of clustering in protein contact maps for alpha helix pairs. In *ICANNGA* (1), 758-766 (2007)
- [15] Grassberger, P.: Pruned-enriched Rosenbluth method: Simulations of θ polymers of chain length up to 1 000 000. *Phys. Rev. E* **56**, 3682-3693 (1997); Frauenkron, H., Bastolla, U., Gerstner, E., Grassberger, P., Nadler, W.: New Monte Carlo algorithm for protein folding.

- Phys. Rev. Lett. **80**, 3149-3152 (1998); Bastolla, U., Frauenkron, H., Gerstner, E., Grassberger, P., Nadler, W.: Testing a new Monte Carlo algorithm for protein folding. *Proteins* **32**, 52-66 (1998); Hsu, H.-P., Mehra, V., Nadler, W., Grassberger, P.: Growth-based optimization algorithm for lattice heteropolymers. *Phys. Rev. E* **68**, 021113 (2003); Growth algorithms for lattice heteropolymers at low temperatures. *J. Chem. Phys.* **118**, 444-451 (2003)
- [16] Hansmann, U., Okamoto, Y.: New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* **9**, 177-183 (1999)
- [17] Iba, Y., Chikenji, G., Kikuchi, M.: Simulation of lattice polymers with multi-self-overlap ensemble. *J. Phys. Soc. Jpn.* **67**, 3327-3330 (1998); Chikenji, G., Kikuchi, M., Iba, Y.: Multi-self-overlap ensemble for protein folding: Ground state search and thermodynamics. *Phys. Rev. Lett.* **83**, 1886-1889 (1999)
- [18] Janke, W.: Multicanonical Monte Carlo simulations. *Physica A* **254**, 164-178 (1998)
- [19] Kauzmann, W.: Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1-63 (1959); Dill, K.A.: The meaning of hydrophobicity. *Science* **250**, 297-298 (1990)
- [20] Kou, S.C., Oh, J., Wong, W.H.: A study of density of states and ground states in hydrophobic-hydrophilic protein folding models by equi-energy sampling. *J. Chem. Phys.* **124**, 244903 (2006)
- [21] Landau, D.P., Binder, K.: *A Guide to Monte Carlo Methods in Statistical Physics*. Cambridge University Press, Cambridge (2000)
- [22] Landau, D.P., Tsai, S.-H., Exler, M.: A new approach to Monte Carlo simulation: Wang-Landau sampling. *Am. J. Phys.* **72**, 1294-1301 (2004)
- [23] Lesh, N., Mitzenmacher, M., Whitesides, S.: A complete and effective move set for simplified protein folding. In *RECOMB*, 188-195 (2003)
- [24] Li, Y.W., Wüst, T., Landau, D.P.: Monte Carlo simulations of the HP model (the “Ising model” of protein folding). *Comput. Phys. Commun.* **182**, 1896-1899 (2011)
- [25] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.M., Teller, E.: Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087-1092 (1953)
- [26] Parsons, D.F., Williams, D.R.M.: An off-lattice Wang-Landau study of the coil-globule and melting transitions of a flexible homopolymer. *J. Chem. Phys.* **124**, 221103 (2006); Paul, W., Strauch, T., Rampf, F., Binder, K.: Unexpectedly normal phase behavior of single ho-

- mopolymer chains. Phys. Rev. E **75**, 060801(R) (2007); Vogel, T., Bachmann, M., Janke, W.: Freezing and collapse of flexible polymers on regular lattices in three dimensions. Phys. Rev. E **76**, 061803 (2007); Seaton, D.T., Wüst, T., Landau, D.P.: A Wang-Landau study of the phase transitions in a flexible homopolymer. Comput. Phys. Commun. **180**, 587-589 (2008); Seaton, D.T., Wüst, T., Landau, D.P.: Collapse transitions in a flexible homopolymer chain: Application of the Wang-Landau algorithm. Phys. Rev. E, **81**, 011802 (2010)
- [27] Prellberg, T., Krawczyk, J.: Flat histogram version of the pruned and enriched Rosenbluth method. Phys. Rev. Lett. **92**, 120602 (2004); Prellberg, T., Krawczyk J., Rechnitzer, A.: Polymer simulations with a flat histogram stochastic growth algorithm. In: Landau, D.P, Lewis, S.P., Schüttler, H.-B. (eds.) Computer Simulation Studies in Condensed-Matter Physics XVII, pp. 122-135. Springer, Berlin Heidelberg New York (2006)
- [28] Rampf, F., Binder, K., Paul, W.: The phase diagram of a single polymer chain: New insights from a new simulation method. J. Polym. Sci. Pol. Phys. **44**, 2542-2555 (2006)
- [29] Sokal, A.D.: Monte Carlo methods for the self-avoiding walk. In: Binder K. (ed.) Monte Carlo and Molecular Dynamics Simulations in Polymer Science, pp. 47-124. Oxford University Press, New York (1995)
- [30] Swetnam, A.D., Allen, M.P.: Improved simulations of lattice peptide adsorption. Phys. Chem. Chem. Phys. **11**, 2046-2055 (2009)
- [31] Wang, F., Landau, D.P.: Efficient, multiple-range random walk algorithm to calculate the density of states. Phys. Rev. Lett. **86**, 2050-2053 (2001); Wang, F., Landau, D.P.: Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. Phys. Rev. E **64**, 056101 (2001); Wang, F., Landau, D.P.: Determining the density of states for classical statistical models by a flat-histogram random walk. Comput. Phys. Commun. **147**, 674-677 (2002)
- [32] Wüst, T., Landau, D.P.: The HP model of protein folding: A challenging testing ground for Wang-Landau sampling. Comput. Phys. Commun. **179**, 124-127 (2008); Versatile approach to access the low temperature thermodynamics of lattice polymers and proteins. Phys. Rev. Lett. **102**, 178101 (2009)
- [33] Zhang, J., Kou, S.C., Liu, J.S.: Biopolymer structure simulation and optimization via fragment regrowth Monte Carlo. J. Chem. Phys. **126**, 225101 (2007)

- [34] Zhou, C., Bhatt, R.N.: Understanding and improving the Wang-Landau algorithm. Phys. Rev. E. **72**, 025701(R) (2005); Liang, F.: A generalized Wang-Landau algorithm for Monte Carlo computation. J. Am. Stat. Assoc. **100**, 1311 (2005)